Summary of DDWG sub-group discussion 2009-01-28

High-level DMAC architecture:

- Federated architecture comprising a system of several Data Assembly Centers (DACs).

(a cartoon of the architecture implied by this discussion might be a helpful communications tool …)

  - Federated architecture is midway between the extremes of fully centralized (a single DAC that takes in all the observations) and fully distributed (every data producer is responsible for assembling and serving their own data in standardized fashion).
  - Examples of federated data management systems abound in ocean observations – GHRSST, Argo, and OceanSites, for example. The Air Traffic Control System (ATCS), with local airports part of a region and regions part of the national airspace is another example, though with a more rigid hierarchal governance structure than is suitable for DMAC.
  - The organization of DACs may be based upon data types or may be based upon Regional governance on a case by case basis. For example, NDBC may be the DAC for all mooring data, whereas outputs from regionally-operated modeling centers may be managed by Regional DACs.
  - The role of a DAC is to provide integrated access to data from all the sources under its purview.
  - "Sources" can include *in situ* sensors, networks of sensors, satellites, numerical models, and other DACs.
  - Every IOOS DAC shall provide access to the data that it holds through standardized interfaces (service types, data encoding) mandated by IOOS. IOOS shall mandate a small number of specific interfaces for different classes of data. In order to foster evolution of DMAC over time it will be commonplace that competing strategies for the same data type may co-exist as designated IOOS standards, with translation technologies ensuring inter-operability between them.
  - The internals of a DAC can be a "black box" for IOOS purposes--only the external interface(s) are of concern. The various strategies by which DACs handle distributed data internally can serve as incubators for the development of future DMAC technologies.
  - Users can access data from any DAC. Users only interested in a localized, near-real-time data may wish to access the DAC closes to the source. Users interested in a larger view from multiple sources or in delayed-mode quality-controlled data may wish to access a larger DAC.

Governance:

- The interagency IOOS program is responsible for specifying (adopting, adapting, developing) the necessary standards.
- Governance process must allow for evaluation and possible inclusion of other possible conventions.

Levels of compliance:

- Details of DAC responsibilities remain tbd -- Is it enough to implement standard interface(s)? Are there also responsibilities to ensure adequate metadata? Also responsibilities to ensure that data are reliably migrated to archives?
- May need levels of compliance. See CMM levels after regional workshop.

Data archiving:

- Not all regional data that goes to NDBC is currently sent on the archive.
- Problem is mainly lack of manpower to do complete quality control.
- It is vital that observations not be lost. It would be better to send data to the archive as-is with a “Quality unknown” flag than not to send data at all.

Gridded Data:

Structured Grid (uniform, rectilinear and curvilinear)

- Mechanisms (data model standards, APIs) are relatively well-developed.

- Use of OpenDAP+CF+NetCDF package is recommended by DMAC-ST.
- Many clients and services exist for regular or rectilinear grid, dramatically fewer for curvilinear grid.
- Look at what GHRSST level 3 and 4 has done and adopt as much as possible. (GHRSST level 2 data may be describable as curvilinear grids with the CF standard and associated tools.)
- Exploration of WCS as a future option should also continue (currently only delivers simple regular grids).

Unstructured Grid (e.g. grids composed of connected triangles)

- Need standard data model, as well as APIs to deliver unstructured grid data via OpenDAP.
- IOOS should play a role in developing infrastructure for unstructured grid model data, since unstructured grid models are the most common models used for inundation, and are increasingly being used in the regions.

Feature collection data (e.g., in situ obs):

- partially developed.
- Implementation and evaluation of SOS should be continued.
- Implementation and evaluation of OPeNDAP-based solutions (e.g. ERDDAP) should likewise continue in parallel.
- In view of the role of DACs above there should be a shift towards greater emphasis on the delivery of data to the applications and software development environments that users depend upon. This mandates the development of data models and client code libraries with well-specified APIs.

Metadata

- We need metadata standards (e.g., profiles of SensorML) for sensors, platforms, and networks thereof, and for the QC processes performed on observations. IOOS should partner with QARTODS and OGC SWE to advance this effort.
- We also need metadata standards for numerical models. IOOS should partner with and support the groups that are already engaged in this activity (e.g Earth System Curator).